



A Chemical Screen for Biological Small Molecule-RNA Conjugates Reveals CoA-Linked RNA

Citation

Kowtoniuk, Walter. E., Yinghua Shen, Jennifer M. Heemstra, Isha Agarwal, and David R. Liu. 2009. A chemical screen for biological small molecule-RNA conjugates reveals CoA-linked RNA. *Proceedings of the National Academy of Sciences of the United States of America* 106(19): 7768-7773.

Published Version

doi:10.1073/pnas.0900528106; <http://dx.doi.org/10.1073/pnas.0900528106>

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:3425857>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

A Chemical Screen for Biological Small Molecule-RNA Conjugates Reveals CoA-Linked RNA

Walter E. Kowtoniuk, Yinghua Shen, Jennifer M. Heemstra, Isha Agarwal, David R. Liu*

*Department of Chemistry and Chemical Biology
12 Oxford Street
Harvard University
Cambridge, Massachusetts 02138*

*Correspondence:

David R. Liu
12 Oxford Street
Cambridge, Massachusetts 02138
drliu@fas.harvard.edu
Tel: (617) 496-1067
Fax: (617) 496-5688

Abstract

In contrast with the rapidly expanding set of known biological roles for RNA, the known chemical diversity of cellular RNA has remained limited primarily to canonical RNA, 3'-aminoacylated tRNAs, nucleobase-modified RNAs, and 5'-capped mRNAs in eukaryotes. We developed two methods to detect in a broad manner chemically labile cellular small molecule-RNA conjugates. The methods were validated by the detection of known tRNA and rRNA modifications. The first method analyzes small molecules cleaved from RNA by base or nucleophile treatment. Application to *E. coli* and *S. venezuelae* RNA revealed an RNA-linked hydroxyfuranone or succinyl ester group, in addition to a number of other putative small molecule-RNA conjugates not previously reported. The second method analyzes nuclease-generated mononucleotides before and after treatment with base or nucleophile and also revealed a number of new putative small molecule-RNA conjugates, including 3'-dephospho-CoA and its succinyl-, acetyl-, and methylmalonyl-thioester derivatives. Subsequent experiments established that these CoA species are attached to *E. coli* and *S. venezuelae* RNA at the 5' terminus. CoA-linked RNA cannot be generated through aberrant transcriptional initiation by *E. coli* RNA polymerase *in vitro*, and CoA-linked RNA in *E. coli* is only found among smaller (< ~200 nucleotide) RNAs that have yet to be identified. These results provide new examples of small molecule-RNA conjugates and suggest that the chemical diversity of cellular RNA may be greater than previously understood.

Introduction

Over the past few decades, RNA has emerged as much more than an intermediary in biology's central dogma. Ribozymes (1), riboswitches (2), microRNAs (miRNAs) (3), small interfering RNAs (siRNAs) (4), Piwi-interacting RNAs (piRNAs) (5), small nuclear RNAs (snRNAs) (6), CRISPR sRNAs (7), RNA transcriptional regulators (8), and long non-coding RNAs (9, 10) are all examples of RNAs that are thought to play a wide range of catalytic, regulatory, or defensive roles in the cell. Models of early biotic systems have proposed even broader roles for RNA, including the possibility that RNA-tethered molecules participated in RNA-templated chemical reactions as an early form of metabolism (11-16).

In contrast with these newer insights into its functional diversity, the known chemical diversity of natural RNA has remained limited primarily to canonical polyribonucleotides, 3'-aminoacylated tRNAs (17), modified nucleobases in a variety of RNAs (18), and 5'-capped mRNAs in eukaryotes (19-21). This disparity between functional and chemical diversity, coupled with the powerful functional properties of synthetic small molecule-nucleic acid conjugates (21-24) led us to speculate that small molecule-RNA conjugates beyond those previously described may exist in modern cells as evolutionary fossils or even as novel RNAs with functions enabled by their modifications.

To begin to explore this possibility, we have developed and implemented a general approach to discovering small molecule-RNA conjugates that is not dependent on a specific type of small-molecule structure or a particular biological function of the conjugate. Our method uses simple chemical reactions on RNA to liberate small-molecule groups or small molecule-conjugated nucleotides. Comparative high-resolution liquid chromatography and mass spectrometry (LC/MS) of these species identifies the masses of labile small molecules or small molecule-nucleotide conjugates that are putatively linked to cellular RNA. MS/MS fragmentation, isotope labeling, and comparison with authentic standards is then used to elucidate the structures of small molecules derived from conjugates with biological RNAs.

Using this approach, we have identified coenzyme A (CoA) and several CoA thioesters as covalent conjugates to cellular RNA in *Escherichia coli* and *Streptomyces venezuelae*. Experiments with *E. coli* RNA polymerase *in vitro* suggest that the observed CoA groups are not installed through aberrant, non-specific transcriptional initiation. In addition, experiments indicate that the CoA-derived RNA(s) are under ~200 nucleotides in length. While the identity of the corresponding RNA(s) and their possible biological relevance are not yet known, our

findings collectively suggest that the chemical diversity of biological RNA in modern cells is greater than previously understood.

Results

Small-Molecule Cleavage Method Detects Known RNA Modifications. We subjected whole cellular RNA from *E. coli* or *S. venezuelae* to size-exclusion chromatography and retained the macromolecular fraction ($> \sim 2,500$ Da). One half of the resulting material was treated with mild aqueous base (pH 8.0) or with a simple alkyl amine nucleophile (500 mM *n*-butylamine in acetonitrile) in order to cleave base-labile and nucleophile-labile small molecules, respectively. The other half was subjected to control conditions (pH 4.5, or acetonitrile with no *n*-butylamine, respectively) designed to leave small molecule-RNA conjugates intact (Fig. 1).

Each sample was separately subjected to size-exclusion chromatography as before, but this time the small-molecule fraction ($< \sim 2,500$ Da) was retained. The two samples were then analyzed by LC/MS. Peaks with corresponding retention times containing species with similar mass:charge ratios (m/z) from the two samples were computationally paired, and their relative abundances were calculated (25). Species that were more abundant in the base- or nucleophile-treated sample relative to the control sample were considered candidate small molecules cleaved from cellular macromolecules (Fig. 1). This comparative analysis proved essential since each separate sample contained thousands of detectable chemical species.

To account for the possible presence of contaminating non-RNA macromolecules in our RNA preparations, we also pre-treated a third whole cellular RNA sample with a mixture of RNase A and RNase T1 before the first size-exclusion step to confirm that candidate small molecules arose from a small molecule-RNA conjugate rather than from small molecules conjugated to other macromolecules. The ion abundance of a genuine small molecule-RNA conjugate, but not that of a contaminating small molecule-macromolecule conjugate, should decrease significantly in samples pretreated with RNase. Small molecules that were more abundant upon treatment with base or nucleophile, but that were less abundant when pretreated with RNase, were considered candidate small molecules cleaved from cellular RNA.

The small-molecule cleavage method was validated *in vitro* using aminoacylated tRNAs as positive controls. Phenylalanine-charged tRNA^{Phe} was prepared *in vitro* from purified phenylalanine-tRNA aminoacyl synthetase (PheRS) and *E. coli* tRNA (26). Comparative high-resolution LC/MS analysis revealed a 65-fold greater abundance of phenylalanine arising from base cleavage conditions compared with the control conditions (Fig. 2A). Similarly, 52-fold and

42-fold ratios of amino acid abundance in base-treated versus control samples were observed with tRNA charged *in vitro* with LeuRS and AspRS, respectively. These results demonstrate that the small-molecule cleavage method is able to detect aminoacylated tRNAs generated *in vitro*.

Next we validated the small-molecule cleavage method by detecting amino acids conjugated to endogenous cellular RNA. Freshly isolated RNA from *E. coli* was subjected to the small-molecule cleavage method (Fig. 2B). To narrow the resulting list of candidates, we defined three criteria: (i) ≥ 4.5 -fold enrichment in the base-treated samples relative to the control samples; (ii) ≥ 3.0 -fold enrichment of a corresponding butyl amine addition product in the nucleophile-treated samples relative to their control samples; and (iii) ≥ 2 -fold lower base enrichment values upon pretreatment with RNases A and T1. These thresholds were empirically determined to be sufficiently low to enable detection of most of the amino acid positive controls, while sufficiently high to avoid most false positives such as canonical mononucleotides predicted to be stable to the base and nucleophiles used.

When this approach was applied to *E. coli* and *S. venezuelae* RNA, 14 of the 20 amino acids (70%) were found to meet all three of the above criteria (Fig. S4). The successful detection of the majority of the amino acids conjugated to RNA validates the ability of the small-molecule cleavage method to detect the presence of known small molecule-RNA conjugates from whole cellular RNA.

Small-Molecule Cleavage Method Detects Unknown RNA Conjugates. In addition to the expected amino acids and known labile nucleobase modifications (data not shown), the small-molecule cleavage method applied to *E. coli* RNA also reveals five unknown species that met all three of the above criteria for putative base- or nucleophile-labile small molecule-RNA conjugates (Fig. S5). When *S. venezuelae* RNA was subjected to the same treatment, 14 amino acids and five unknown species were found to meet the same three criteria. The smallest of these unknown species, with $[M+H]^+$ $m/z = 101.0232$, was found in both *E. coli* and *S. venezuelae*. Independent biological replicates of total *E. coli* RNA subjected to this method generated consistent enrichment factors with trial-to-trial correlation coefficients of ~ 0.85 ; similarly, two independent trials of the small-molecule cleavage method applied to *S. venezuelae* RNA produced datasets with a correlation coefficient of 0.87.

A library of synthetic RNA 45-mers of random sequence was subjected to the complete RNA isolation and small-molecule cleavage method described above. None of the putative small

molecule-RNA conjugates arising from the analysis of *E. coli* or *S. venezuelae* RNA were significantly enriched when using synthetic RNA. These results indicate that the putative small-molecule RNA conjugates from bacterial RNA arise from cellular processes, and not from RNA degradation or rearrangement reactions that occur during the small-molecule cleavage method.

Structural Elucidation of $m/z = 101.0232$. We chose the $[M+H]^+$ $m/z = 101.0232$ species (Figs. 2C and S6) as our initial target for structural elucidation. The observed mass and isotope profile of the base-cleaved species suggested a molecular formula of $C_4H_4O_3$ (expected $[M+H]^+$ $m/z = 101.0233$). A corresponding *n*-butylamine-treated cleavage product was observed with $[M+H]^+$ $m/z = 156.1013$, suggesting that this unidentified species could arise from an RNA-linked oxyester or thioester that undergoes hydrolysis in the presence of mild base to form a carboxylic acid, and aminolysis in the presence of *n*-butylamine to form an *n*-butylamide.

In addition, analysis of the *n*-butylamine-treated RNA from both *E. coli* and *S. venezuelae* revealed the presence of a species consistent with the double addition of *n*-butylamine to the $C_4H_4O_3$ unknown and the loss of two molecules of water (observed $[M+H]^+$ $m/z = 211.1826$; expected m/z for $C_{12}H_{23}N_2O = 211.1805$; also see Supporting Information). Excluding ketenes, allenes, allene oxides, and oxocyclopropanes due to their rarity among biological molecules, we proposed seven possible carboxylic acids consistent with the molecular formula $C_4H_4O_3$, of which only three (molecules **1**, **2** and **3**) are capable of undergoing a second *n*-butylamine addition with loss of water (Fig. 2D).

Authentic samples of the *n*-butyl amides of **1**, **2**, and **3** (*n*-butyl amides **9**, **10**, and **11**, respectively) were prepared by chemical synthesis (Figs. S1-3). The LC/MS spectra of the cellular *n*-butyl amide of $[M+H]^+$ $m/z = 101.0232$ did not match those of synthetic *n*-butyl amides **9** or **10**, indicating that the unknown is not ketone **1** or *trans*-alkene **2** (Fig. 2E). The remaining candidate, compound **3**, preferentially exists as the hydroxyfuranone **4**, which can spontaneously tautomerize in aqueous solution to form succinic anhydride (27, 28). The *n*-butyl amide of **4** (compound **11**) was synthesized and found to spontaneously isomerize to *n*-butyl succinimide (**12**). LC/MS analysis revealed that **12** matched the *n*-butyl amide of the cellular unknown. Importantly, the MS/MS ion fragmentation patterns of **12** and the *n*-butyl amide of the cellular unknown were virtually identical (Fig. 2F). Taken together, these results are consistent with a model in which the observed 100.0154 Da base-cleaved species is hydroxyfuranone **4** or its tautomer, succinic anhydride (Fig. S7).

Subsequent experiments did not reveal any RNA nucleotides directly conjugated to a 100.0154 Da small molecule (*vide infra*). Therefore, we hypothesized that **4** is not directly conjugated to RNA, but instead arises from a base- and nucleophile-labile succinyl group of a larger small molecule-RNA conjugate, the identity of which we sought to reveal using a method capable of detecting small molecule-linked nucleotides.

Nucleotide Cleavage Method Detects Known Small Molecule-Nucleotide Conjugates. The small-molecule cleavage method described above can reveal labile small molecules that are directly or indirectly conjugated to RNA, but does not characterize intact small molecule-nucleotide conjugates as they might exist in cellular RNA. We therefore developed a complementary method to identify nucleotides conjugated to base-labile or nucleophile-labile small molecules. In this second method, the macromolecular fraction of whole cellular RNA was treated with nuclease P1, an endonuclease that cleaves RNA to generate mononucleotides with a 3' hydroxyl group and a 5' phosphate (29). As in the first method, one half of the resulting sample was treated with base (pH 10.5) or nucleophile (500 mM *n*-butylamine in acetonitrile) while the other half was treated with control conditions (pH 4.5, or acetonitrile, respectively). The samples were subjected to size-exclusion chromatography again and the small-molecule fraction from each was retained. Following comparative high-resolution LC/MS, species with greater abundance in control samples versus the base- or nucleophile-treated samples were considered candidate nucleotides linked to labile small molecules (Fig. S8).

As with the small-molecule cleavage method, the nucleotide cleavage method was also validated by detection of amino acid-linked RNAs from whole cellular *E. coli* and *S. venezuelae* RNA. An enrichment threshold of ≥ 2 -fold was empirically found to distinguish the 3'-aminoacyl adenosine monophosphates, which serve as base-labile positive controls, from known nucleotide modifications such as *N*⁶,*N*⁶-dimethyladenosine that should not be base labile. Species enriched ≥ 2 -fold included 15-16 of the 20 major 3'-aminoacyl adenosine monophosphates (Fig. S9), as well as many species consistent with rRNA and tRNA nucleoside modifications (Fig. S10) that have been previously reported, although not necessarily known to exist in *E. coli* or *S. venezuelae* (18). These results validate the ability of the nucleotide cleavage method to detect the presence of known small molecule-RNA conjugates.

Nucleotide Cleavage Method Detects Unknown Small Molecule-Nucleotide Conjugates. In addition to 3'-aminoacyl adenosine monophosphates and known nucleotide modifications, 17

unknown species were enriched ≥ 2 -fold in *E. coli* nucleotide samples treated with control conditions relative to base-treated samples (Fig. S11). For *S. venezuelae* RNA, the nucleotide cleavage method revealed 18 unknown species that were enriched 2-fold or more (Fig. S11). Independent biological replicates of the nucleotide cleavage method generate enrichment factors with trial-to-trial correlation coefficients of 0.93-0.95 (Figs. 3 and S12). None of the observed unknown species were detected from total *E. coli* or *S. venezuelae* RNA if nuclease P1 was omitted, or if nuclease P1 treatment was replaced with incubation in formamide and/or 10 mM EDTA at 95°C, conditions expected to impair RNA secondary structure. These results suggest that the species in Fig. 3 arise from nuclease P1-mediated RNA cleavage, and not from the liberation of small molecules non-covalently bound to RNA.

Structural Elucidation of $m/z = 786.1582$ and 686.1432 . One unknown species from both *E. coli* and *S. venezuelae* detected by the nucleotide cleavage method was $[M-H]^-$ $m/z = 786.1582$ (Figs. 4A and S13). We became especially interested in this species because its MS/MS spectrum always included a major fragment with $[M-H]^-$ $m/z = 686.1209$ (Fig. 4B), and because a second unknown species with a similar observed mass ($[M-H]^-$ $m/z = 686.1432$) was slightly *more* abundant upon treatment with base in both bacteria. We hypothesized that these two species might represent a larger, base- and nucleophile-labile small molecule-nucleotide conjugate (787.1660 Da), and a smaller version (687.1510 Da) that is left behind after the loss of a 100.0154 Da $C_4H_4O_3$ moiety such as the succinyl group hypothesized above to be conjugated to RNA. In support of this model, the MS/MS fragmentation daughter ions of the $[M-H]^-$ $m/z = 686.1432$ species represented a subset of the daughter ions arising from fragmentation of the $[M-H]^-$ $m/z = 786.1582$ ion (Figs. S14 and S15).

The molecular weights of these two unknowns are too large to unambiguously assign empirical formulas. We therefore cultured *S. venezuelae* in media containing ^{13}C -glucose as the sole carbon source, or in media containing ^{15}N -ammonium sulfate as the sole nitrogen source. Total *S. venezuelae* RNA from each culture was separately treated with nuclease P1 and analyzed by LC/MS. The resulting shifts in observed m/z values allowed us to assign a molecular formula of $C_{25}H_{38}N_7O_{16}P_2S$ to the compound with $[M-H]^-$ $m/z = 786.1582$ and a molecular formula of $C_{21}H_{34}N_7O_{13}P_2S$ to the compound with $[M-H]^-$ $m/z = 686.1432$ (Fig. S16). The smaller compound therefore represents the loss of $C_4H_4O_3$ from the larger, consistent with the above model.

Inspection of the MS/MS fragmentation patterns of both unknowns strongly suggested that both species contain ADP. Therefore, we reasoned that the 787.1660 Da species likely consists of a 100.0154 Da group and a 260.1211 Da group attached to the pyrophosphate of ADP (Figs. 4, S14, and S15). The fragmentation data also suggested that the 687.1510 Da species is the same 260.1211 Da group attached to the pyrophosphate of ADP, but lacking the 100.0154 Da group. Collectively, these observations led us to propose that the 786.1582 species and 686.1432 species are 3'-dephospho-succinyl-CoA (expected $[M-H]^- m/z = 786.1577$) and 3'-desphospho-CoA (expected $[M-H]^- m/z = 686.1416$), respectively. These hypotheses were confirmed by LC/MS comparison of the cellular species with authentic 3'-dephospho-succinyl-CoA and authentic 3'-dephospho-CoA (Figs. 4C).

A search for additional related CoA derivatives in our data sets together with LC/MS analysis of authentic standards revealed that RNA from *E. coli* and *S. venezuelae* both contain 3'-dephospho-acetyl-CoA (observed $[M-H]^- m/z = 728.1532$; expected $[M-H]^- m/z = 728.1522$), and that RNA from *S. venezuelae* contains 3'-dephospho-methylmalonyl-CoA (observed $[M-H]^- m/z = 786.1585$; expected $[M-H]^- m/z = 786.1577$). The absence of a 3'-dephospho-methylmalonyl-CoA signal from *E. coli* RNA is consistent with the known inability of *E. coli* to biosynthesize methylmalonyl-CoA, assuming that biosynthesis of methylmalonyl-CoA precedes covalent attachment to RNA.

During the base-induced cleavage of small molecules from RNA, the succinyl group of succinyl-CoA-RNA can undergo cyclization to generate succinic anhydride while cleaving itself from the CoA-RNA. Indeed, analysis of either succinic anhydride or succinic acid under the LC/MS conditions used for the small-molecule cleavage method resulted in a LC peak with identical $[M+H]^+ m/z$, retention time, and MS/MS fragmentation pattern as that of the cellular 100.0154 Da species (Fig. S7). Together, these findings suggest that the $[M+H]^+ m/z = 101.0232$ ion discovered through the small-molecule cleavage method is derived from the base-induced cleavage of the succinyl group from succinyl-CoA-RNA.

Characterization of the Attachment of CoA Derivatives to RNA. CoA and its thioester derivatives are common cellular metabolites (30). To ensure that the detected CoA species did not simply arise from the 3' phosphatase activity of nuclease P1 on intracellular CoA and CoA esters that had unexpectedly survived RNA purification and size exclusion, we spiked varying quantities of CoA thioesters into *E. coli* and *S. venezuelae* cell lysates, and repeated the RNA isolation, nuclease P1 digestion, and LC/MS analysis. Despite adding up to 10,000-fold more

acetyl-CoA and succinyl-CoA than we observed in unspiked samples, no significant changes in the abundance of the corresponding 3'-dephosphorylated species were detected (Fig. 4D). Likewise, the addition to cell lysates of similarly large quantities of benzoyl-CoA, butyryl-CoA, and crotonyl-CoA, three CoA esters that were not observed in our original experiments, did not result in corresponding detectable levels of those compounds (Fig. 4D). These results demonstrate that the CoA species observed in our experiments on *E. coli* and *S. venezuelae* RNA cannot be accounted for by endogenous small-molecule contaminants, and further support the conclusion that these species arise from cellular small molecule-RNA conjugates.

Based on the structure of 3'-dephospho-CoA, we hypothesized that these modifications are present at the 5' termini of one or more cellular RNAs. To test this hypothesis we digested total RNA from both *E. coli* and *S. venezuelae* with nuclease P1 in the presence of ^{18}O -enriched water. Because nuclease P1 catalyzes the attack of a water molecule on RNA to generate 5'-phosphonucleotides (29), in the presence of ^{18}O water all nuclease P1 digestion products other than the nucleotides at the 5' termini will have a mass shift of +2 Da compared with products of digestion in ^{16}O water. Indeed, the expected +2 Da shift was observed for 3'-Phe-AMP (observed $[\text{M-H}]^- m/z = 495.1290$; expected $[\text{M-H}]^- m/z = 495.1285$). In contrast, no mass shift from nuclease P1 digestion in the presence of ^{18}O water was observed for any of the 3'-dephospho-CoA derivatives, consistent with a model in which these species are originally present at the 5' termini of RNA molecules (Fig. S17).

By comparing signal intensities of cellular and authentic samples of known concentration, we estimate that there are 80-120 total copies of CoA-RNA and CoA-thioester-RNA per *E. coli* or *S. venezuelae* cell. The amount of total CoA-linked per μg of total RNA is ~ 8 femtomoles for *E. coli* and ~ 13 femtomoles for *S. venezuelae*.

Transcriptional Initiation by *E. coli* RNA Polymerase In Vitro Cannot Account for Observed Levels of CoA-RNA. Since 3'-dephospho-CoA shares structural features with ATP and is a known biosynthetic precursor of CoA (30), we speculated that CoA might be incorporated into RNA at the 5'-terminus through aberrant transcriptional initiation with 3'-dephospho-CoA or its thioesters instead of ATP. Indeed, this mechanism of CoA incorporation into a transcript *in vitro* has been reported with T7 RNA polymerase (31). To explore this possibility we carried out *in vitro* transcription using *E. coli* RNA polymerase in the presence of high concentrations of 3'-dephospho-CoA using two different templates. For the first template, we modified a pUC19 plasmid to encode an adenosine at the +1 position of each of its four

predicted transcripts. An *in vitro* transcription reaction containing 0.5 mM of each NTP and either 0.5 mM or 5 mM of 3'-dephospho-CoA yielded 555 µg or 544 µg of RNA, respectively. When this RNA was purified, digested with nuclease P1, and analyzed by LC/MS, no 3'-dephospho-CoA was detected. The second template used was *E. coli* genomic DNA. *In vitro* transcription in the presence of either 0.5 mM or 5 mM of 3'-dephospho-CoA yielded 89 µg or 95 µg of RNA. Once again, this material contained no detectable 3'-dephospho-CoA after nuclease P1 digestion (Fig. S18). In contrast, when a 5'-CoA-linked transcript (generated using T7 RNA polymerase) was spiked into an *in vitro* transcription reaction and processed in the same way, CoA-linked RNA was readily detected (Fig. S18).

Based on the observed abundances of CoA-RNA from *E. coli* cells, we would expect to obtain more than 2.8 pmol of 3'-dephospho-CoA from ~550 µg of A-initiated RNA, and 0.45 pmol of 3'-dephospho-CoA from ~90 µg of RNA transcribed from the *E. coli* genome, if aberrant transcriptional initiation were predominantly responsible for the CoA-RNA conjugates. These quantities should be readily detected by our methods, which can reliably detect ≤ 0.1 pmol of 3'-dephospho-CoA (Fig. S18). If one assumes that the inability of *E. coli* RNA polymerase to incorporate these levels of 3'-dephospho-CoA *in vitro* reflects an inability to do so *in vivo*, these results suggest that CoA groups are installed post-transcriptionally.

Size Distribution of CoA-Linked RNAs. Both methods described above subject RNA to size exclusion to remove molecules of molecular weight $< \sim 2,500$ Da. To establish an upper size limit on CoA-linked RNAs, we subjected the macromolecule fraction to further size fractionation using silica-based RNA purification columns (Qiagen RNeasy columns), which separate RNA molecules into two fractions that are less than or greater than ~200 nucleotides in length (Fig. S19). Each of the two fractions was then subjected to nuclease P1 digestion and LC/MS analysis.

As expected, 3'-aminoacyl adenosine monophosphates conjugated to tRNAs (~76 nucleotides) were present predominantly in the < 200 nucleotide flow-through fraction (Fig. 9B), and the rRNA nucleoside modification $N^6,N^{6'}$ -dimethyladenine (conjugated to 1.5 kDa-2.9 kDa rRNAs) was detected in the > 200 base fraction (Fig. 4E) (32, 33). Like the 3'-aminoacyl adenosine monophosphates, the CoA-linked nucleotides were predominantly detected in the flow-through RNA fraction. This result suggests that the CoA-linked RNA(s) from *E. coli* and *S. venezuelae* are not widely distributed in their size but instead are below ~200 nucleotides in length. In addition, this finding further supports the hypothesis that the CoA modifications arise

through a mechanism other than non-specific transcriptional initiation, which would be expected to generate a broad size distribution of CoA-linked RNAs.

Discussion

We have developed and validated two methods that in principle enable the detection of any base- or nucleophile-labile small molecule-RNA conjugate. Application of these methods led to the discovery of a hydroxyfuranone or succinyl group, as well as a series of CoA derivatives including succinyl-CoA, linked to *E. coli* and *S. venezuelae* RNA. These findings represent new examples of biological small molecule-RNA conjugates beyond aminoacylated tRNAs, RNAs containing modified nucleobases, and 5'-capped mRNA in eukaryotes. More generally, our results suggest that the chemical diversity of cellular RNA is greater than previously understood. Since *E. coli* and *S. venezuelae* represent two different phyla, our findings suggest that the presence of these newly discovered conjugates is not limited to a narrow range of species.

The 3'-dephospho-CoA group is attached to the 5' terminus of cellular RNA(s) of length < ~200 nucleotides. On average we observe ~100 CoA-RNA molecules per *E. coli* cell, which suggests that CoA-linked RNAs together are approximately ten-fold less abundant than Phe-linked tRNA in *E. coli* (34) and ~10-100-fold less abundant than the *E. coli* 6S RNA (36). While we currently do not know the biological role, if any, that these CoA-RNA conjugates may play, it is tempting to speculate that they might play a role in RNA stability, RNA localization, or gene regulation, or even in mediating chemical reactions involving CoA groups linked to RNA strands that serve to direct reactivity (21, 24). The last possibility highlights an unusual feature of these groups compared with most previously discovered RNA modifications—namely, that CoA and CoA thioesters are substantially more reactive.

From *E. coli* RNA we observe 3'-dephospho-CoA, succinyl-dephospho-CoA, and acetyl-dephospho-CoA as RNA conjugates. In addition, we observe methylmalonyl-dephospho-CoA as a RNA conjugate from *S. venezuelae*. These observations suggest that CoA attachment to RNA occurs after thioesterification. The liberation of 3'-dephospho-CoA derivatives from cellular RNA by nuclease P1 digestion together with their presence on the 5' terminus of RNA (Fig. S17) strongly suggests that the CoA-RNA linkage is a phosphodiester bond linking the 3' phosphate of CoA to the 5' end of the RNA.

Although our *in vitro* transcription experiments suggest that non-specific transcriptional initiation is not the primary mechanism for CoA-RNA formation, they do not exclude the

possibility of a gene-specific transcriptional initiation pathway for CoA incorporation, or even a non-specific transcriptional pathway if other cellular components beyond those present in the *in vitro* transcription reactions are required. DNA primase synthesizes short RNAs that prime DNA synthesis (37) and in theory could also serve as possible source of CoA-linked RNAs. Primase-generated RNAs have been observed to be ten-fold less abundant in *E. coli* (38) than CoA-linked RNAs, however, arguing against this possibility. Studies are ongoing to identify additional small molecule-RNA conjugates, to characterize the RNA species to which these groups are attached, and to evaluate their possible functional roles in the cell.

Experimental Methods

See the Supporting Information for additional experimental details.

Small-Molecule Cleavage Method. 1 mg of *E. coli* RNA or 750 µg of *S. venezuelae* RNA as prepared above was subjected to cleavage conditions (base: 500 mM NH₄HCO₃, pH 8.0, 37 °C, 2.5 hrs; nucleophile: 500 mM *n*-butylamine in acetonitrile, 37 °C, 8.0 hrs). An equal quantity of RNA was subjected to control conditions (base control: 500 mM NH₄OAc, pH 4.5, 37 °C, 2.5 hrs; nucleophile control: acetonitrile, 300 µL, 37 °C, 8.0 hrs). After treatment, the samples were acidified with 200 µL of 3 M NH₄OAc, pH 4.5. The small-molecule fraction was isolated by size-exclusion chromatography using a NAP5 column and lyophilized. The lyophilized product was redissolved in 20 µL of 0.1% aqueous sodium formate and analyzed by LC/MS. For the experiment using synthetic RNA, 2.3 µmol of a library of random synthetic N₄₅ RNAs (IDT) was dissolved in cell lysis buffer, and processed as described above.

LC/MS Data Collection and Analysis. LC/MS was performed using a Waters Aquity UPLC Q-TOF Premier instrument with an Aquity UPLC BEH C18 column. See the Supporting Information for a detailed description of LC/MS and MS/MS conditions. The analysis of total ion chromatograms was performed using the XCMS program (25). Integrated ion abundances were averaged among replicates, and the ratios of these average ion intensities between cleavage conditions and control conditions were the enrichment values reported

Base-cleaved species were matched with corresponding nucleophile-cleaved species by subtracting 55.07858 ± 0.020 Da (the mass of butylamine minus the mass of water) from the *n*-butyl amide cleavage products. For the purpose of this study, base-cleaved species without a

corresponding nucleophile-cleaved partner were discarded, even though some small molecule-RNA conjugates were overlooked as a result.

Nucleotide Cleavage Method. 350 μg of *E. coli* RNA or 250 μg of *S. venezuelae* RNA was digested with 10 U nuclease P1 (Sigma-Aldrich) in 200 μL of 50 mM NH_4OAc , pH 4.5 at 37 °C for 20 min). The digestion products were purified by size-exclusion chromatography (NAP5) and the small-molecule fraction was retained. Half of the resulting nucleotides were subjected to cleavage conditions (base: 500 mM $(\text{NH}_4)_2\text{CO}_3$, pH 10.5, 37 °C, 2.5 hrs; nucleophile: 500 mM *n*-butylamine in acetonitrile, 37 °C, 8.0 hrs) while the other half was subjected to control conditions (base control: 500 mM NH_4OAc , pH 4.5, 37 °C, 2.5 hrs; nucleophile control: nucleophile control: acetonitrile, 300 μL , 37 °C, 8.0 hrs). The samples were acidified with 200 μL of 3 M NH_4OAc , pH 4.5, lyophilized, redissolved in 20 μL of 0.1% aqueous ammonium formate, and analyzed by LC/MS. The nuclease P1 digestion with H_2^{18}O (Cambridge Isotope Laboratories) was performed as described above except in buffer with a final composition containing 86% H_2^{18}O and 14% H_2^{16}O .

Acknowledgments

This work was supported by the Howard Hughes Medical Institute and the NIH/NIGMS (R01GM065865). We thank Jack Szostak and Matt Hartman for aminoacyl-tRNA synthetase enzymes. W.E.K gratefully acknowledges an NSF Graduate Research Fellowship.

References

1. Doudna JA & Cech TR (2002) The chemical repertoire of natural ribozymes. *Nature* 418(6894):222-228.
2. Mandal M & Breaker RR (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* 5(6):451-463.
3. Chen K & Rajewsky N (2007) The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8(2):93-103.
4. Matzke MA & Birchler JA (2005) RNAi-mediated pathways in the nucleus. *Nat Rev Genet* 6(1):24-35.
5. Brower-Toland B, *et al.* (2007) Drosophila PIWI associates with chromatin and interacts directly with HP1a. *Genes Dev* 21(18):2300-2311.
6. Patel SB & Bellini M (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucl Acids Res* 36(20):6482-6493.
7. Sorek R, Kunin V, & Hugenholtz P (2008) CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* 6(3):181-186.
8. Storz G, Altuvia S, & Wassarman KM (2005) An abundance of RNA regulators. *Annu Rev Biochem* 74(1):199-217.
9. Dinger ME, *et al.* (2009) NRED: a database of long noncoding RNA expression. *Nucl Acids Res* 37:D122-126.
10. Mattick JS & Makunin IV (2006) Non-coding RNA. *Hum Mol Genet* 15:R17-29.
11. Illangasekare M & Yarus M (1999) Specific, rapid synthesis of Phe-RNA by RNA. *Proc Natl Acad Sci USA* 96(10):5470-5475.
12. Szostak JW, Bartel DP, & Luisi PL (2001) Synthesizing life. *Nature* 409(6818):387-390.
13. Benner SA, Ellington AD, & Tauer A (1989) Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci USA* 86(18):7054-7058.
14. Jeffares DC, Poole AM, & Penny D (1998) Relics from the RNA world. *J Mol Evol* 46(1):18-36.
15. Visser CM & Kellogg RM (1978) Bioorganic chemistry and the origin of life. *J Mol Evol* 11(2):163-169.
16. White HB (1976) Coenzymes as fossils of an earlier metabolic state. *J Mol Evol* 7(2):101-104.
17. Hoagland MB, Stephenson ML, Scott JF, Hecht LI, & Zamecnik PC (1958) A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* 231(1):241-257.
18. Dunin-Horkawicz S, *et al.* (2006) MODOMICS: a database of RNA modification pathways. *Nucl Acids Res* 34:D145-149.
19. Wei CM & Moss B (1975) Methylated nucleotides block 5'-terminus of vaccinia virus messenger RNA. *Proc Natl Acad Sci USA* 72(1):318-322.
20. Furuichi Y & Miura K-I (1975) A blocked structure at the 5' terminus of mRNA from cytoplasmic polyhedrosis virus. *Nature* 253(5490):374-375.
21. Li X & Liu DR (2004) DNA-templated organic synthesis: nature's strategy for controlling chemical reactivity applied to synthetic molecules. *Angew Chem, Int Ed* 43(37):4848-4870.
22. Kanan MW, Rozenman MM, Sakurai K, Snyder TM, & Liu DR (2004) Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* 431(7008):545-549.
23. Gartner ZJ, *et al.* (2004) DNA-templated organic synthesis and selection of a library of macrocycles. *Science* 305(5690):1601-1605.
24. Gartner ZJ & Liu DR (2001) The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J Am Chem Soc* 123(28):6961-6963.

25. Smith CA, Want EJ, O'Maille G, Abagyan R, & Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* 78(3):779-787.
26. Hartman MCT, Josephson K, & Szostak JW (2006) Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci USA* 103(12):4356-4361.
27. Poskonin VV & Badovskaya LA (2003) Unusual conversion of 5-hydroxy-2(5H)furanone in aqueous solution. *Chem Heterocycl Compd* 39(5):594-597.
28. Skrin·rov· Z, Bowden K, & Fabian WMF (2000) An ab initio and density functional study on the ring-chain tautomerism of (Z)-3-formyl-acrylic acid. *Chem Phys Lett* 316(5-6):531-535.
29. Romier C, Dominguez R, Lahm A, Dahl O, & Suck D (1998) Recognition of single-stranded DNA by nuclease P1: high resolution crystal structures of complexes with substrate analogs. *Proteins: Struct, Funct, Bioinf* 32(4):414-424.
30. Leonardi R, Zhang Y-M, Rock CO, & Jackowski S (2005) Coenzyme A: back in action. *Prog Lipid Res* 44(2-3):125-153.
31. Huang F (2003) Efficient incorporation of CoA, NAD and FAD into RNA by in vitro transcription. *Nucl Acids Res* 31(3):e8.
32. Brosius J, Palmer ML, Kennedy PJ, & Noller HF (1978) Complete nucleotide sequence of a 16S ribosomal RNA gene from Escherichia coli. *Proc Natl Acad Sci USA* 75(10):4801-4805.
33. Brosius J, Dull TJ, & Noller HF (1980) Complete nucleotide sequence of a 23S ribosomal RNA gene from Escherichia coli. *Proc Natl Acad Sci USA* 77(1):201-204.
34. Jakubowski H & Goldman E (1984) Quantities of individual aminoacyl-tRNA families and their turnover in Escherichia coli. *J Bacteriol* 158(3):769-776.
35. Starr JL, Fefferman R, & with the technical assistance of Sara LE (1964) The occurrence of methylated bases in ribosomal ribonucleic acid of Escherichia coli K12 W-6. *J Biol Chem* 239(10):3457-3461.
36. Wassarman KM & Storz G (2000) 6S RNA Regulates E. coli RNA Polymerase Activity. *Cell* 101(6):613-623.
37. Frick DN & Richardson CC (2001) DNA PRIMASES. *Annu Rev Biochem* 70(1):39.
38. Ogawa T, Hirose S, Okazaki T, & Okazaki R (1977) Mechanism of DNA chain growth: XVI. Analyses of RNA-linked DNA pieces in Escherichia coli with polynucleotide kinase. *J Mol Biol* 112(1):121-140.

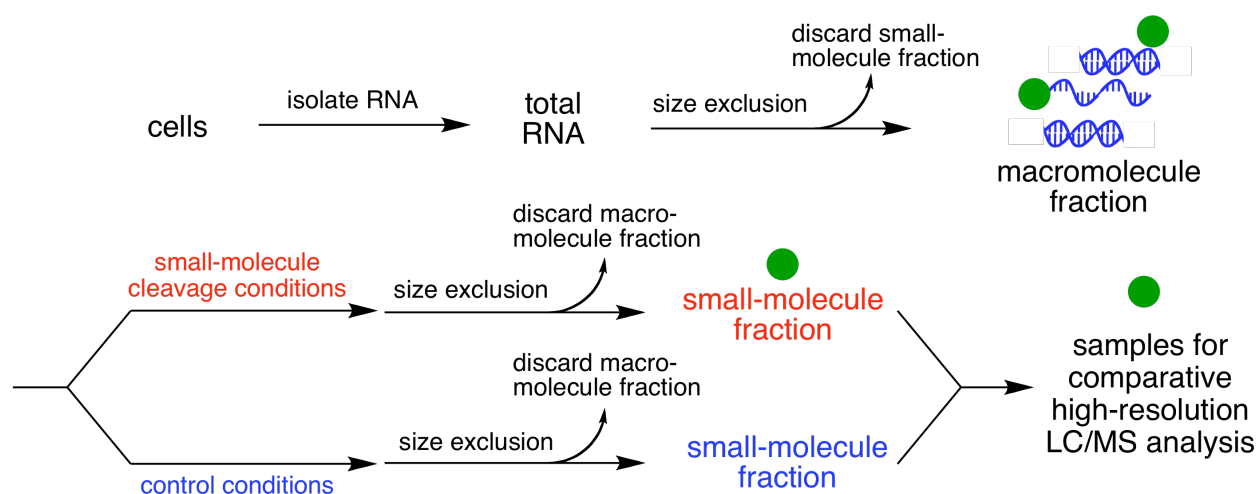


Fig. 1. Small-molecule cleavage method for small molecule-RNA conjugate discovery.

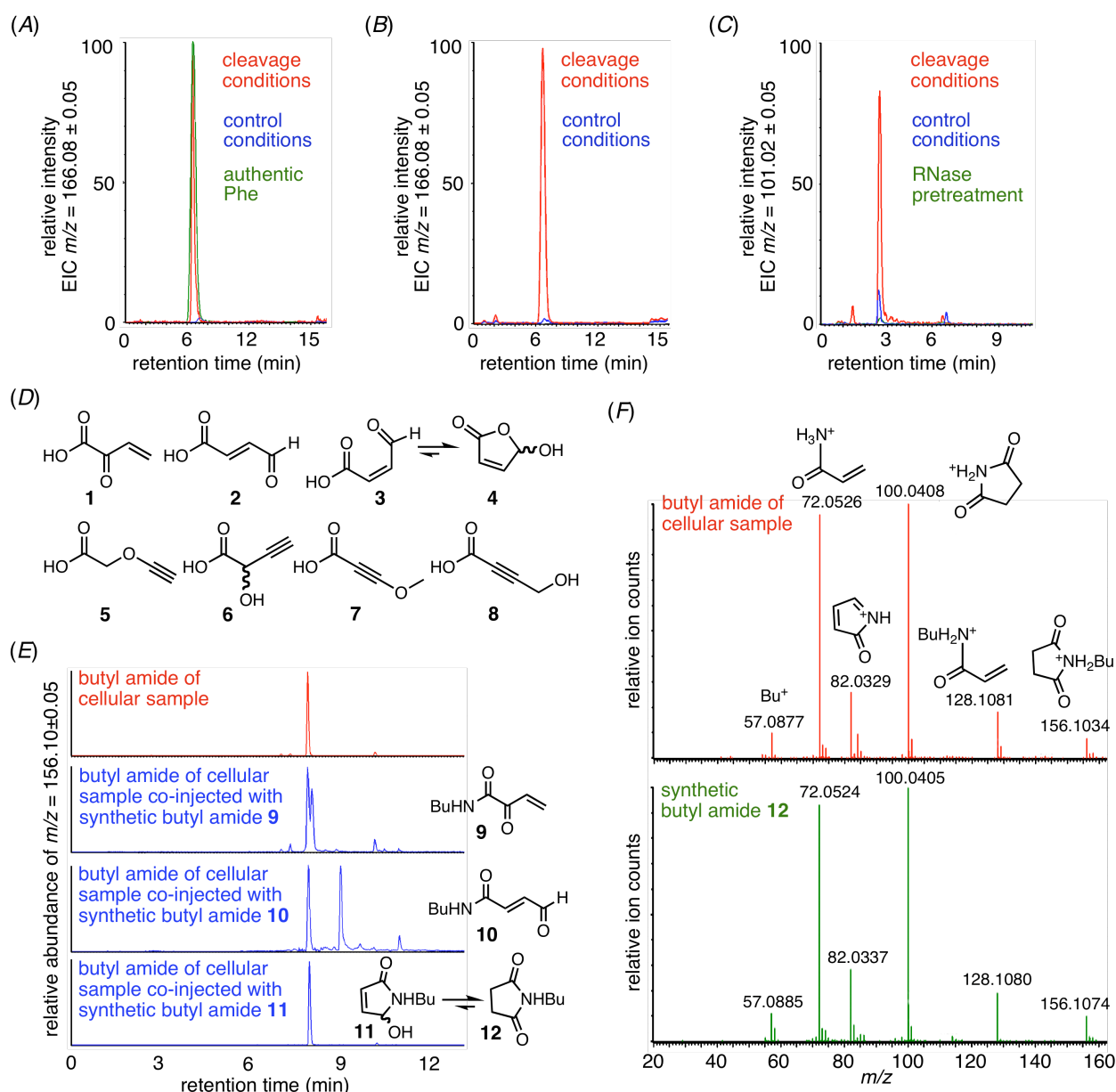


Fig. 2. Initial application of the small-molecule cleavage method and structural elucidation of $[M+H]^+ m/z = 101.023$. (A) Purified *E. coli* tRNA was aminoacylated *in vitro* with phenylalanine and PheRS, then subjected to the small-molecule cleavage method and analyzed by LC/MS. The extracted ion chromatogram (EIC) at $[M+H]^+ m/z = 166.08$ (corresponding to phenylalanine) for both samples is shown here; the cleavage conditions (pH 8.0) result in 65-fold higher Phe abundance than the control conditions (pH 4.5). (B) EIC for the experiment in (A) using total RNA isolated from *E. coli* instead of aminoacylated tRNA. (C) The EICs for the unknown ion of $[M+H]^+ m/z = 101.0232$ from *E. coli* RNA subjected to pH 8.0 cleavage conditions, pH 4.5 control conditions, or pre-treatment with RNase A and RNase T1 prior to pH 8.0 cleavage conditions. (D) Possible carboxylic acids of the formula $C_4H_4O_3$, excluding ketenes, allenes, allene oxides, and oxocyclopropanes. (E) Co-injection of the *n*-butyl amide variants of candidates 1, 2, or 3 (compounds 9, 10, or 11, respectively) with the cellular butyl amide reveals that the cellular butyl amide matches hydroxyfuranone butyl amide 11. (F) MS/MS fragmentation of cellular *n*-butyl amide (top) and synthetic 12 (bottom) confirms that the $[M+H]^+ m/z = 101.023$ species is the hydroxyfuranone 4 or its aqueous tautomers.

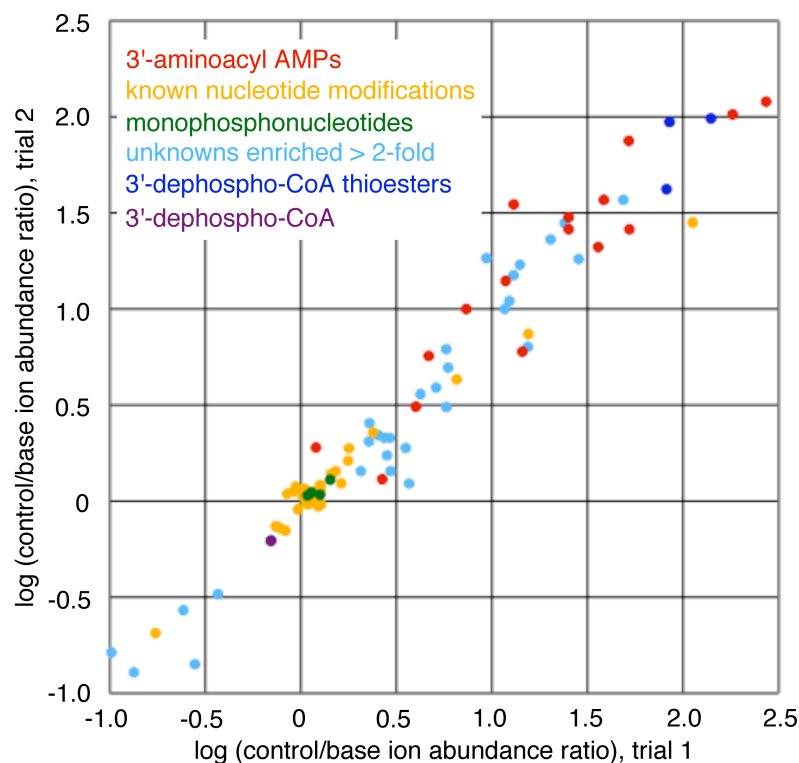


Fig. 3. Result of two independent trials ($r = 0.95$) of the nucleotide-cleavage method applied to total *S. venezuelae* RNA. The observed species include sixteen 3'-aminoacyl adenosine monophosphates, 16 known nucleotide modifications, the four canonical RNA nucleotides, 3'-dephospho-CoA and its three thioester derivatives discovered in this work, and 18 additional unknown species with a control:base ratio ≥ 2 -fold. Note that 3'-dephospho-CoA is observed with an control:base ratio < 1 due to the base-labile nature of the 3'-dephospho-CoA thioesters.

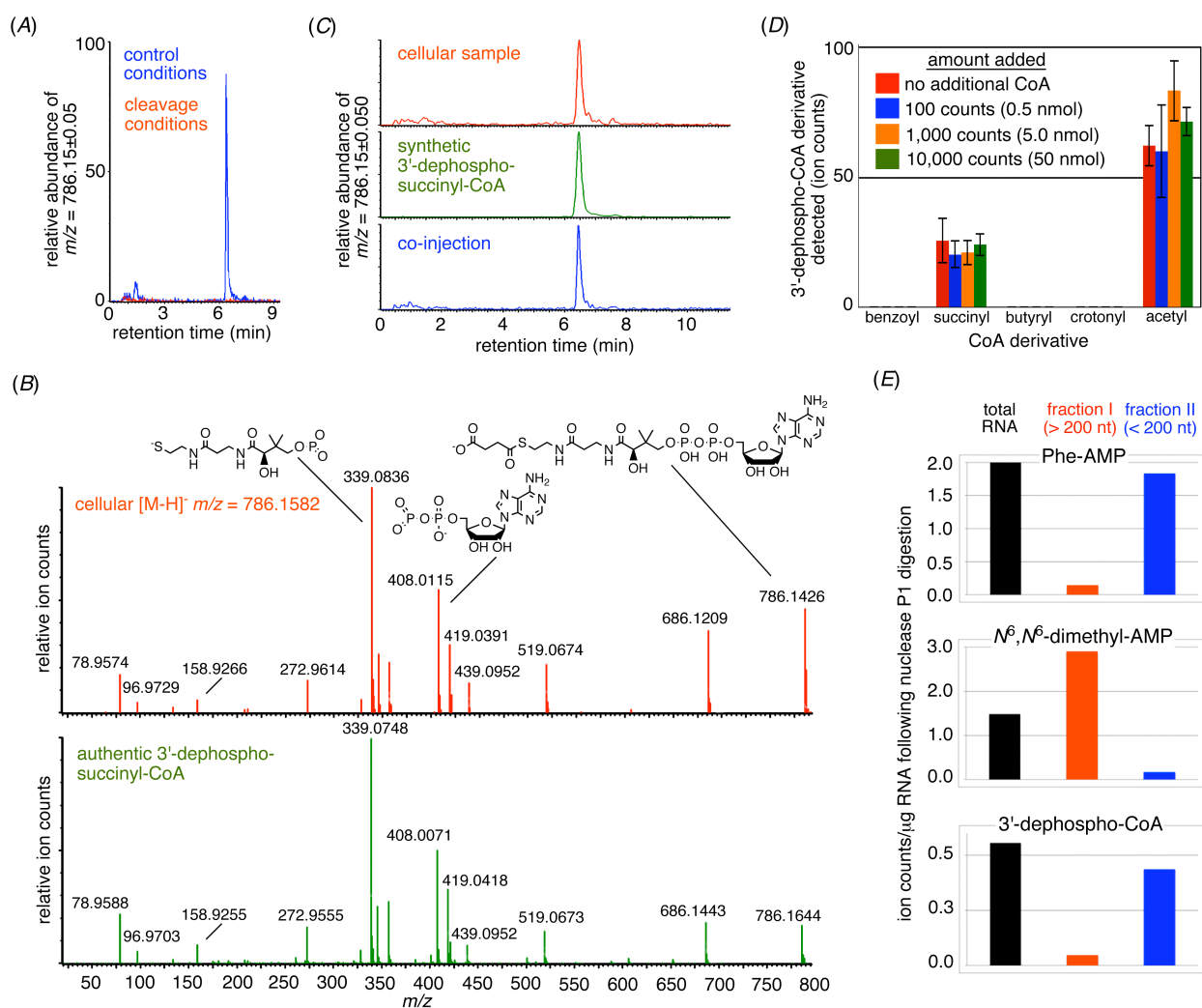


Fig. 4. Two small molecule-linked nucleotides of $[M-H]^-$ $m/z = 786.1432$ and 686.1432 from *E. coli* and *S. venezuelae* RNA. (A) The EICs for $[M-H]^-$ $m/z = 786.1532$ from *E. coli* RNA digested with nuclease P1 and subjected to cleavage conditions (pH 10.5) or control conditions (pH 4.5). (B) MS/MS fragmentation of the $[M-H]^-$ $m/z = 786.1582$ species from *E. coli* and of authentic 3'-dephospho-succinyl-CoA. See Figs. S14 and S15 for a plausible complete fragment assignment. (C) EIC comparison of the *E. coli* cellular RNA nuclease P1 digest and authentic 3'-dephospho-succinyl-CoA. (D) Spiking large quantities of CoA thioesters into *E. coli* cell lysate prior to RNA isolation and the nucleotide cleavage method does not change the observed ion counts of these species, indicating that the observed 3'-dephospho-CoA signals do not arise from small-molecule CoA thioester contaminants. (E) Total *E. coli* RNA was separated into RNAs of length $> \sim 200$ nucleotides (fraction I) and RNAs of length $< \sim 200$ nucleotides (fraction II) using a silica column (Qiagen RNeasy). Each fraction was subjected to nuclease P1 digestion and analyzed by LC/MS. The presence of 3'-dephospho-CoA in fraction II suggests that the CoA-linked RNA(s) are primarily $< \sim 200$ nucleotides in length.